New insights into the correlation structure of DSM-IV depression symptoms in the general population v. subsamples of depressed individuals

Foster, Simon ; Mohler-Kuo, Meichun

Abstract: AIMS: Previous research failed to uncover a replicable dimensional structure underlying the symptoms of depression. We aimed to examine two neglected methodological issues in this research: (a) adjusting symptom correlations for overall depression severity; and (b) analysing general population samples v. subsamples of currently depressed individuals. METHODS: Using population-based cross-sectional and longitudinal data from two nations (Switzerland, 5883 young men; USA, 2174 young men and 2244 young women) we assessed the dimensions of the nine DSM-IV depression symptoms in young adults. In each general-population sample and each subsample of currently depressed participants, we conducted a standardised process of three analytical steps, based on exploratory and confirmatory factor and bifactor analysis, to reveal any replicable dimensional structure underlying symptom correlations while controlling for overall depression severity. RESULTS: We found no evidence of a replicable dimensional structure across samples when adjusting symptom correlations for overall depression severity. In the general-population samples, symptoms correlated strongly and a single dimension of depression severity was revealed. Among depressed participants, symptom correlations were surprisingly weak and no replicable dimensions were identified, regardless of severity-adjustment. CONCLUSIONS: First, caution is warranted when considering studies assessing dimensions of depression because general population-based studies and studies of depressed individuals generate different data that can lead to different conclusions. This problem likely generalises to other models based on the symptoms' inter-relationships such as network models. Second, whereas the overall severity aligns individuals on a continuum of disorder intensity that allows non-affected individuals to be distinguished from affected individuals, the clinical evaluation and treatment of depressed individuals should focus directly on each individual's symptom profile.

DOI: https://doi.org/10.1017/S2045796016001086

New insights into the correlation structure of DSM-IV depression symptoms in the general population versus subsamples of depressed individuals

S. Foster

M. Mohler-Kuo

Running head

Dimensions of depression symptoms

Authors' affiliation:

Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Zurich, Switzerland

Corresponding author:

Simon Foster

Epidemiology, Biostatistics and Prevention Institute

University of Zurich

Hirschengraben 84

8001 Zürich, Switzerland

Phone: +41 44 634 46 22

Fax: +41 44 634 49 86

Email: simon.foster@uzh.ch

Word count: 3997

Abstract

Aims: Previous research failed to uncover a replicable dimensional structure underlying the symptoms of depression. We aimed to examine two neglected methodological issues in this research: a) adjusting symptom correlations for overall depression severity; and b) analysing general population samples versus subsamples of currently-depressed individuals.

Methods: Using population-based cross-sectional and longitudinal data from two nations (Switzerland, 5883 young men, USA, 2174 young men and 2244 young women) we assessed the dimensions of the nine DSM-IV depression symptoms in young adults. In each general-population sample and each subsample of currently-depressed participants, we conducted a standardized process of three analytical steps, based on exploratory and confirmatory factor and bifactor analysis, to reveal any replicable dimensional structure underlying symptom correlations while controlling for overall depression severity.

Results: We found no evidence of a replicable dimensional structure across samples when adjusting symptom correlations for overall depression severity. In the general-population samples, symptoms correlated strongly and a single dimension of depression severity was revealed. Among depressed participants, symptom correlations were surprisingly weak and no replicable dimensions were identified, regardless of severity-adjustment.

Conclusions: First, caution is warranted when considering studies assessing dimensions of depression because general population-based studies and studies of depressed individuals generate different data that can lead to different conclusions. This problem likely generalizes to other models based on the symptoms' inter-relationships such as network models. Second, whereas the overall severity aligns individuals on a continuum of disorder intensity that allows non-affected individuals to be distinguished from affected individuals, the clinical evaluation and treatment of depressed individuals should focus directly on each individual's symptom profile.

Keywords:

Depressive disorder; factor analysis; epidemiology; diagnosis; classification

Introduction

Major depression, which is characterized by the core symptoms of depressed mood and anhedonia (World Health Organization, 2012, American Psychiatric Association, 2013), is a leading contributor to the global burden of disease (Murray & Lopez, 1996, Bromet *et al.*, 2011, Ferrari *et al.*, 2013, Kessler & Bromet, 2013, Cuijpers *et al.*, 2014). For years, debate has raged over the clinical presentation of depression (Baumeister & Parker, 2012, van Loo *et al.*, 2012). Specifically, studies have been conducted seeking to identify symptom-based dimensions and subtypes via statistical analysis of symptom co-occurrence using factor analysis, principal component analysis and latent class analysis (Chen *et al.*, 2000, Aggen *et al.*, 2005, Shafer, 2006, Carragher *et al.*, 2009, Aggen *et al.*, 2011, Cole *et al.*, 2011, Mezuk & Kendler, 2012, Hybels *et al.*, 2013, Buhler *et al.*, 2014, Li *et al.*, 2014a, b, Rodgers *et al.*, 2014, Fried *et al.*, 2016). Tightly-correlated symptom sets are important as they might constitute dimensions or subtypes that imply different aetiologies and/or treatment responses. However, as summarized in a recent systematic review, these studies have failed to generate replicable results (van Loo *et al.*, 2012).

Two methodological issues have not yet been considered, however. First, previous studies failed to disentangle two distinct sources of correlation between any two symptoms. Such correlations may be 1) due to differences in overall severity (i.e., individuals with more severe depression score higher for all symptoms than individuals with less severe depression; hence, symptoms A and B are correlated); or 2) due to a specific profile of symptom correlations (e.g., individuals scoring high for symptom A could typically score high for symptom B too, but not necessarily for symptom C which is more closely linked to symptom D). Therefore, it may be more appropriate to study symptom correlations adjusted for overall depression severity. If a structure underlying the symptoms exists, it should be revealed more clearly in this way.

Second, most studies were conducted on samples of depressed individuals. Studies that examined the dimensions of depression in general population samples, however, consistently revealed one single dimension of depression severity (Muthén, 1989, Aggen *et al.*, 2005, Aggen *et al.*, 2011, Cole *et al.*, 2011, Mezuk & Kendler, 2012, Familiar *et al.*, 2015). Thus, in the general population, depression was found to be a uni-dimensional construct. Apparently, it makes a difference whether one studies the general population or depressed individuals only, but this issue remained unaddressed.

The present study's main objective was to examine the dimensional structure of the nine symptoms of depression listed in the DSM-IV a) when adjusting symptom correlations for overall depression severity; and b) in general-population samples versus subsamples of currently-depressed individuals. We adopted a dimensional approach, since evidence increasingly suggests that many psychiatric syndromes, including depression, are continuous and hence dimensional rather than categorical (Slade & Andrews, 2005, Goldberg *et al.*, 2009, Prisciandaro & Roberts, 2009, Markon *et al.*, 2011, Haslam *et al.*, 2012, Eaton *et al.*, 2013).

Materials and methods

Study design

We used a) longitudinal data from the Cohort Study on Substance Use Risk Factors (C-SURF); and b) cross-sectional data from the U.S. National Health and Nutrition Survey (NHANES). In total, we considered four samples: C-SURF baseline, C-SURF follow-up, NHANES men, and NHANES women. Comparing C-SURF baseline and follow-up data permitted us to examine whether results were replicable across two time points in the same

sample. Comparing the C-SURF and NHANES data allowed us to examine whether results were replicable in two different populations.

Each of the four samples was analyzed twice: once in the full (general-population) sample, and once only in the subsample of participants within a current mild-to-severe depressive episode, generating eight analytical samples in total.

Participants

C-SURF is a large cohort study examining young men in Switzerland, for which details on sampling and non-response bias have been published elsewhere (Studer *et al.*, 2013a, Studer *et al.*, 2013b). It was designed to be representative of young non-institutionalized Swiss men. The study protocol was approved by the Ethics Committee for Clinical Research at Lausanne University Medical School (protocol number 15/07), and all subjects consented to participate.

5990 men completed the baseline survey between September 2010 and March 2012. Of these, 107 (1.8%) were excluded for missing data on the depression items. Of these, 5155 (87.6%) answered all necessary items of the follow-up survey performed between January 2012 and April 2013. The mean time elapsed between baseline and follow-up was 1.3 years (standard deviation 0.2).

NHANES is a continuous cross-sectional survey released in 2-year cycles (Centers for Disease Control and Prevention: National Center for Health Statistics). It was designed to be representative of the non-institutionalized U.S. civilian population. NHANES study protocols were approved by the National Center for Health Statistics Research Ethics Review Board, and all participants consented.

We included NHANES data from NHANES cycles 2005-2012 for men and women from 18-28 years old, an age range chosen to resemble the C-SURF cohort. Men and women were analyzed separately. Of the total 2371 men and 2542 women, 197 (8.3%) and 298 (11.7%) were excluded for missing depression items data.

Measures

C-SURF: Self-reported depressive symptoms were assessed via the Major Depressive Inventory – WHO-MDI (Bech *et al.*, 2001, Olsen *et al.*, 2003). This validated measure covers DSM-IV and ICD-10 depression symptoms over the past 14 days, using 12 items with 6-point answer scales ranging from "never" (0) to "all the time" (5). Items were aggregated into the nine DSM-IV symptoms, as proposed previously (Bech *et al.*, 2001) (Table 1). Subjects were classified as having 'no', 'mild', 'moderate', or 'severe' depression based on the MDI summation score (Olsen *et al.*, 2003). For correlation and factor analyses, symptoms were dichotomized into present/absent, as per ICD-10 definitions (Bech *et al.*, 2001) (Table 1).

NHANES 2005-2012: Self-reported depressive symptoms were assessed via the Patient Health Questionnaire (PHQ-9) (Kroenke *et al.*, 2001). This validated measure covers the nine DSM-IV depression symptoms over the past 14 days (Table 1). Four answer options are provided, ranging from "not at all" (0) to "nearly every day" (3). Participants were classified as having 'no', 'mild', 'moderate', or 'severe' depression based on the PHQ9 summation score (Kroenke *et al.*, 2001). Note that the threshold score we used to denote 'mild' depression was termed 'moderate' depression by Kroenke et al. This threshold resembled most closely the threshold for 'mild' depression that we used in the C-SURF sample, in terms of the percentage summation score required for the diagnosis (40% in C-SURF, 37% in NHANES). For correlation and factor analyses, symptoms were dichotomized into present/absent, as per DSM-IV definitions (Kroenke *et al.*, 2001) (Table 1).

Statistical analysis

First, we examined tetrachoric correlations of the depression symptoms. To compare the correlations of each general-population sample with those of the corresponding subsample of currently-depressed subjects, we calculated the ratio of the squared correlations for each symptom pair and used Steiger's test to formally examine the hypothesis that the two correlation matrices differed (Steiger, 1980). Steiger's test sums the squared differences of the Fisher transformed correlations of the two matrices and tests this sum against the chi-square distribution. Second, we assessed the dimensionality of the depression symptoms in three steps, each step conducted separately for each of the eight samples to determine whether the results were replicable.

Step 1: We first performed one-factor confirmatory factor analysis (CFA) and exploratory factor analysis (EFA). CFA consisted of the nine symptoms as indicators of an underlying depression factor, thereby modelling overall depression severity (model 1). With EFA, we tested one- to seven-factor models to determine which best fit the data. Both CFA and EFA were estimated using mean and variance-adjusted weighted least squares (WLSMV) estimations, which is the standard for categorical indicators (Barendse *et al.*, 2014, Li *et al.*, 2014a). Model fit was evaluated via standard criteria for good model fit (Aggen *et al.*, 2005, Li *et al.*, 2014a): Root Mean Square Error of Approximation (RMSEA) ≤ 0.05; Comparative Fit Index (CFI) ≥ 0.95; and Tucker-Lewis Index (TLI) ≥ 0.95. For EFA, the model with the lowest number of factors achieving these criteria was adopted (model 2).

Step 2: From model 1, we derived the modification indices for the residual symptom co-variances as indicators of symptom pairs correlated beyond the general factor (i.e., as indicators of substantial severity-adjusted symptom correlations). Modification indices

estimate the degree of improvement in model fit if the corresponding parameter is included in the model (Brown & Moore, 2012). Consequently, the modification index of a residual co-variance indicates whether the model would fit better if this co-variance was included in the CFA model.

We considered a modification index ≥ 3.84 statistically significant (Brown & Moore, 2012). We then re-fitted the one-factor CFA, this time including the residual symptom co-variances revealed by the modification indices (model 3). The residual symptom correlations derived from this CFA model generated an estimate of symptom correlations corrected for overall depression severity. If there is a dimensional structure beyond overall depression severity, these correlations (a) should be replicable across the samples, and (b) form interpretable symptom clusters. Because adopting CFA models based on modification indices is associated with a high risk of over-fitting (MacCallum *et al.*, 1992), we used the median of each modification index across 5000 case-based bootstrap samples.

Step 3: Finally, we estimated a series of bifactor models that, by definition, consist of one general factor and several group factors. Each indicator variable loads simultaneously on the general factor and one of the group factors (Reise *et al.*, 2010). Thus, bifactor models allow for estimating group factors controlled for a general factor (Reise *et al.*, 2010) and, hence, correspond directly to our notion of assessing depression dimensions (the group factors) controlled for overall depression severity (the general factor). If there is a replicable dimensional structure underlying the depression symptoms, at least one of the bifactor models should either converge with the residual correlations revealed in step 2, or provide an alternative model that is replicable across samples.

We used two approaches to identify the group factors:

a. We examined three theoretically-derived groupings (models 4.a1-3):
    1. Three genetic factors revealed by Kendler et al. (Kendler *et al.*, 2013).
    2. The common distinction of a cognitive/affective factor versus a somatic factor, as defined in the systematic review by van Loo et al. (van Loo *et al.*, 2012);
    3. The symptoms most consistently found on a single factor in the review (van Loo *et al.*, 2012) versus the remaining symptoms;
b. A non-rotated EFA that comprises several factors can be rotated into a bifactor structure (Jennrich & Bentler, 2011, 2012), resulting in an exploratory bifactor analysis (EBFA) that can then undergo confirmatory analysis. We derived the EBFA from the EFA calculated in step 1 and re-fitted it as a CFA model (model 4.b). However, only one EFA model revealed a sufficient number of factors to be rotated into a bifactor structure. We therefore used this bifactor model across all samples, rather than assessing a separate bifactor solution for each sample.

Bifactor models were estimated using WLSMV estimation and model fit was evaluated as in step 2.

Analyses were performed using R-software version 3.1.2 (R Core Team, 2014), particularly using the packages "psych" (Revelle, 2013), "semTools" (Pornprasertmanit *et al.*, 2013), and "lavaan" (Rosseel, 2012). R-scripts are available at https://osf.io/a6tuw/.

Results

Participants' baseline characteristics are summarized in Table 2. Prevalence rates for current depression of at least mild degree ranged from 4.9% to 7.6%.

Symptom correlations

Substantial symptom correlations were revealed in the general-population samples (median correlations ranging from r = 0.55 to 0.74), while correlations in the depressed samples were surprisingly weak (median correlations from r = 0.04 to 0.24, Table 3). Correlations were greater in the general-population samples, and these differences were pronounced: in average correlations were higher by a factor ranging between 8.4 and 30.9 across samples (Table 4). Steiger tests confirmed that all general-population sample correlation matrices differed significantly from their counterparts in the depressed samples (Table 4). Only one correlation among women ("life not worth living" and "appetite changes") was slightly higher in the depressed sample (ratio of squared correlation = 0.8, Table 3).

Factor analyses

Step 1 revealed that the one-factor model fit the data very well in all general-population samples. This was revealed by both CFAs and EFAs (Table 5, models 1 and 2). In contrast, in depressed samples, no replicable dimensional structure was identified. Specifically, the one-factor CFAs failed to achieve good model fit in three of four samples and the EFAs revealed different numbers of factors across samples.

Step 2 indicated that 50 of 288 (8 samples x 36 symptom pairs) possible residual co-variances (17.4%) were substantial. Including these residual co-variances in the CFAs improved the fit of all models and resulted in good-fitting models, except for the NHANES sample of depressed women (Table 5, model 3). Both positive and negative correlations were revealed, positive correlations ranging from 0.10 to 0.48 (median: r = 0.29) and negative correlations from -0.46 to -0.07 (median: r = -0.26, Table 6). However, the correlations failed to exhibit any replicable pattern across the samples, and 17 of the 50 correlations (34.0%) were not statistically significant (Table 6).

In *Step 3*, no bifactor model was replicable across the samples (Table 5, models 4.a1-4.b). The most stable model was model 4.a2, which achieved good model fit in three of four NHANES samples and one C-SURF sample. Note that all models were inadmissible in at least four of the eight samples, due to negative residual variances.

A closer look at the bifactor models revealed two issues (see supplementary material available at https://osf.io/a6tuw/). First, 18 of 24 models that were inadmissible were inadmissible because at least one of the group factors consisted of one large factor loading, with all other loadings being virtually zero, thereby leading to a model that was empirically under-identified (Kline, 2011). Furthermore, this pattern of one very large loading with otherwise negligible loadings is indicative of over-factoring (i.e., the inclusion of unnecessary factors) (Rindskopf, 1984). In the remaining six inadmissible models, the majority or all of the loadings were non-significant for at least one group factor. Second, among the admissible models, six had at least one group factor with only non-significant factor loadings, and only two models had significant factor loadings across both the general and group factors. Thus, bifactor analysis provided no evidence for any dimensional structure existing beyond the general severity factor.

Discussion

Main findings

We sought to examine the dimensions underlying the nine DSM-IV depressive symptoms in young adults while adjusting symptom correlations for overall depression severity, and while comparing general-population samples versus subsamples of currently-depressed individuals. Analyses revealed three main results. First, adjusting symptom correlations for overall depression severity left little substantial correlation between the symptoms, and we failed to find any evidence to support a replicable dimensional structure when correcting symptom

correlations for overall depression severity. Second, in the general-population samples, symptoms correlated substantially and were uni-dimensional. Third, among depressed individuals, symptom correlations were mostly weak and there was no evidence of any replicable dimensional structure, regardless of whether or not correlations were adjusted for overall severity.

Our finding that depressive symptoms were uni-dimensional in the general population is totally consistent with results from previous studies that analyzed combined samples of healthy and affected individuals. These studies included general population samples from the USA and Mexico, and youths ages 5-18 in the USA and United Kingdom (Muthén, 1989, Aggen *et al.*, 2005, Aggen *et al.*, 2011, Cole *et al.*, 2011, Mezuk & Kendler, 2012, Familiar *et al.*, 2015). Our results replicate these results among young adults in the USA and extend them to young Swiss men. Furthermore, they resemble recent results reported by Fried et al. who found that, as samples of American and Dutch depression patients became more heterogeneous with respect to overall depression severity, average symptom correlations increased and the factor structures became simpler (Fried *et al.*, 2016).

Our finding indicates that, within the general population, depression can be described by a single dimension of severity, the main reason being that depressed individuals form a comparably homogeneous group, relative to the large majority of individuals who are mostly or completely symptom-free (data not shown). The sizeable symptom correlations found in the general population samples mainly reflected this difference between depressed and non-depressed individuals. As such, the common set of ICD-10 and DSM-IV depression symptoms has diagnostic utility identifying individuals suffering from depression within the general population, and the listed symptoms seem to capture the basic scope and severity of the syndrome well.

Conversely, the uni-dimensionality of depression symptoms was not present among depressed individuals and we found no evidence of any replicable dimensional structure. Our failure to uncover such a structure is totally consistent with a recent systematic review that failed to identify any conclusive evidence that data-driven dimensions or subtypes of depression exist (van Loo *et al.*, 2012, see also Chen *et al.*, 2000, Aggen *et al.*, 2005, Shafer, 2006, Carragher *et al.*, 2009, Aggen *et al.*, 2011, Cole *et al.*, 2011, Mezuk & Kendler, 2012, Hybels *et al.*, 2013, Buhler *et al.*, 2014, Li *et al.*, 2014a, b, Rodgers *et al.*, 2014, Fried *et al.*, 2016). Furthermore, the factor structure of depression changes over time among depressed patients (Fried *et al.*, 2016). It therefore seems unlikely that a dimensional structure underlies the symptoms of depression. Consequently, previous literature reporting and making use of depression dimensions should be considered cautiously.

The mostly-weak symptom correlations among depressed individuals were particularly surprising. Symptom correlations have seldom been reported in the literature and, hence, this phenomenon seems to have gone unnoticed. Nonetheless, Cramer et al. reported average symptom correlations among American adults with a 'dysphoric episode' (defined as an episode with at least two depressive symptoms), and Fried et al. reported average symptom correlations among American and Dutch depression patients. Consistent with our results, these authors reported average correlations ranging from r = 0.17 to 0.23 (Cramer *et al.*, 2012) and from r = 0.12 to 0.39 (Fried *et al.*, 2016). Additionally, previous studies failed to detect substantial stability of depression symptoms and subtypes between successive depressive episodes (Coryell *et al.*, 1994, Lewinsohn *et al.*, 2003, Melartin *et al.*, 2004, Oquendo *et al.*, 2004). Thus, symptom correlations seem to be rather weak, both within and between depressive episodes.

That symptom correlations were so weak implies that, even if a replicable dimensional model existed, it would be based on an average correlation of $r \approx 0.20$; the vast majority of symptom variance would remain unexplained, as correlation-based models cannot explain symptom variance beyond these correlations. This agrees with two recent studies that uncovered highly-diverse symptom profiles among depression patients (Fried & Nesse, 2015, Zimmerman *et al.*, 2015). For example, Fried and Nesse identified 1030 unique profiles of depression symptoms in a sample of 3703 depressed American outpatients, with the most frequent profile only occurring in 1.8% of patients (Fried & Nesse, 2015). One explanation of how such diverse profiles develop is that adverse life events and other risk factors exerted differential impacts on depressive symptoms (Keller & Nesse, 2006, Keller *et al.*, 2007, Lux & Kendler, 2010, Fried *et al.*, 2014, Fried *et al.*, 2015) and appeared to change the symptoms' correlation patterns (Cramer *et al.*, 2012). Thus, an individual's symptom profile depends at least partially on the aetiological factors that provoked the depressive episode. Furthermore, these different aetiologies are likely to imply differential responses to various treatment options. For example, evidence indicates that depression related to negative life events and trauma is more responsive to psychotherapy than to medication, whereas depressed individuals with maladaptive personality traits may respond better to selective serotonin-reuptake inhibitors (Simon & Perlis, 2010).

Two final issues concern the recent emergence of network models as an alternative account of mental disorders (Bringmann *et al.*, 2013, Goekoop & Goekoop, 2014, van Borkulo *et al.*, 2014, Boschloo *et al.*, 2015, Bringmann *et al.*, 2015, van Borkulo *et al.*, 2015, Beard *et al.*, 2016). Network models are based on the premise that symptom inter-relationships reflect direct causal influences between symptoms, rather than underlying latent factors, as in the factor analysis framework. The exact relationship between factor and network models remains unclear, however (Molenaar, 2010, Ross, 2010), and various authors disagreed with the

network proponents' critique of the latent variable approach (Belzung *et al.*, 2010, Danks *et al.*, 2010, Haig & Vertue, 2010, Humphry & McGrane, 2010, Markus, 2010). Most importantly, no empirical comparison of these two approaches has yet been reported (Krueger *et al.*, 2010). Thus, how and to what degree one would draw different conclusions when applying factor analysis versus network modelling to one and the same sample is unclear. Future research needs to address this issue.

Second, our results are likely of importance to network research, since they indicate that the choice of sample type can impact the strength of symptom relationships considerably. Since networks are also based on symptom relationships, this should be an issue in network research, too. Indeed, depression-related network studies have been based on all sorts of samples (Bringmann *et al.*, 2013, Goekoop & Goekoop, 2014, van Borkulo *et al.*, 2014, Boschloo *et al.*, 2015, Bringmann *et al.*, 2015, van Borkulo *et al.*, 2015, Beard *et al.*, 2016). Even more intriguing, it was recently found that global network connectivity increased as disorder severity decreased over time (Beard *et al.*, 2016).

Limitations

Our study had several limitations. First, it was restricted to young adults, so the results' generalizability must be re-examined in demographically-broader samples. Second, symptom lists that are more differentiated than the nine DSM-IV criteria might be required, especially considering the weak correlations we detected in our depressed samples. More differentiated symptoms might be needed to capture depression subtypes in patient samples. Note, however, that studies using more comprehensive symptom sets have thus far also failed to uncover replicable dimensions (van Loo *et al.*, 2012). Third, we used dichotomized symptom scores to facilitate comparisons against previous research. Doing so, some information might have been lost. Future studies should evaluate more finely-grained symptom scales. Fourth, step 2 of our

analysis was exploratory and included multiple testing. Note, however, that we used a bootstrap procedure and replicated our analyses across different samples to safeguard against this. Finally, contrary to subtype research using latent class analysis, a dimensional approach could not detect subtypes that are based on only one or two symptoms (if a subtype is defined by several symptoms, however, these symptoms would be tightly correlated and, hence, emerge as a dimension). Thus, whereas our results rule out a dimensional structure of depression, there might still be subtypes of depression characterised by the presence of one or two specific symptoms. Note, however, that previous research focusing on statistically-derived subtypes has also failed to reveal replicable results (van Loo *et al.*, 2012).

Implications

Given prior research findings, our results have two implications. First, caution is warranted when considering studies assessing dimensions of depression because general population-based studies and studies of depressed individuals generate different data that can lead to different conclusions. This problem likely generalizes to other models based on the symptoms' inter-relationships (e.g., network models). Second, it appears that the two dominant aspects of depression are its overall severity and each individual's symptom profile. Whereas the overall severity aligns individuals on a continuum of disorder intensity that allows non-affected individuals to be distinguished from affected individuals, the clinical evaluation and treatment of depressed individuals should focus directly on each individual's symptom profile, since it seems to convey most clinically-relevant information.

Financial support

This work was supported by the Swiss National Science Foundation (grant number FN33CS30_148493). The funder had no role in study design, data collection or analysis, decision to publish, or preparation of the manuscript.

Conflicts of interest

None.

Ethical standards

All procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975 revised in 2008.

Availability of data and materials

The raw data of the C-SURF-cohort study and the NHANES study are available at http://www.c-surf.ch/en/30.html and at http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm.

References

Aggen SH, Kendler KS, Kubarych TS, Neale MC (2011). Differential age and sex effects in the assessment of major depression: a population-based twin item analysis of the DSM criteria. *Twin Research and Human Genetics* **14**, 524-538.

Aggen SH, Neale MC, Kendler KS (2005). DSM criteria for major depression: evaluating symptom patterns using latent-trait item response models. *Psychological Medicine* **35**, 475-487.

American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders 5th edition: DSM-5. American Psychiatric Publishing: Washington, DC.

Barendse MT, Oort FJ, Timmerman ME (2014). Using Exploratory Factor Analysis to Determine the Dimensionality of Discrete Responses. *Structural Equation Modeling: A Multidisciplinary Journal* **22**, 87-101.

Baumeister H, Parker G (2012). Meta-review of depressive subtyping models. *Journal of Affective Disorders* **139**, 126-140.

Beard C, Millner AJ, Forgeard MJ, Fried EI, Hsu KJ, Treadway MT, Leonard CV, Kertz SJ, Bjorgvinsson T (2016). Network analysis of depression and anxiety symptom relationships in a psychiatric sample. *Psychological Medicine* Sep 14 [Epub ahead of print].

Bech P, Rasmussen NA, Olsen LR, Noerholm V, Abildgaard W (2001). The sensitivity and specificity of the Major Depression Inventory, using the Present State Examination as the index of diagnostic validity. *Journal of Affective Disorders* **66**, 159-164.

Belzung C, Billette de Villemeur E, Lemoine M, Camus V (2010). Latent variables and the network perspective. *The Behavioral and Brain Sciences* **33**, 150-151.

Boschloo L, van Borkulo CD, Rhemtulla M, Keyes KM, Borsboom D, Schoevers RA (2015). The Network Structure of Symptoms of the Diagnostic and Statistical Manual of Mental Disorders. *PLoS One* **10**, e0137621.

Bringmann LF, Lemmens LH, Huibers MJ, Borsboom D, Tuerlinckx F (2015). Revealing the dynamic network structure of the Beck Depression Inventory-II. *Psychological Medicine* **45**, 747-757.

Bringmann LF, Vissers N, Wichers M, Geschwind N, Kuppens P, Peeters F, Borsboom D, Tuerlinckx F (2013). A network approach to psychopathology: new insights into clinical longitudinal data. *PLoS One* **8**, e60188.

Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, de Graaf R, Demyttenaere K, Hu C, Iwata N, Karam AN, Kaur J, Kostyuchenko S, Lepine JP, Levinson D, Matschinger H, Mora ME, Browne MO, Posada-Villa J, Viana MC, Williams DR, Kessler RC (2011). Cross-national epidemiology of DSM-IV major depressive episode. *BMC Medicine* **9**, 90.

Brown TA, Moore MT (2012). Confirmatory factor analysis. In Handbook of structural equation modeling (ed. R. H. Hoyle), pp. 361-379. The guilford press: New York.

Buhler J, Seemuller F, Lage D (2014). The predictive power of subgroups: an empirical approach to identify depressive symptom patterns that predict response to treatment. *Journal of Affective Disorders* **163**, 81-87.

Carragher N, Adamson G, Bunting B, McCann S (2009). Subtypes of depression in a nationally representative sample. *Journal of Affective Disorders* **113**, 88-99.

Centers for Disease Control and Prevention: National Center for Health Statistics National Health and Nutrition Examination Survey.

Chen L, Eaton WW, Gallo JJ, Nestadt G (2000). Understanding the heterogeneity of depression through the triad of symptoms, course and risk factors: a longitudinal, population-based study. *Journal of Affective Disorders* **59**, 1-11.

Cole DA, Cai L, Martin NC, Findling RL, Youngstrom EA, Garber J, Curry JF, Hyde JS, Essex MJ, Compas BE, Goodyer IM, Rohde P, Stark KD, Slattery MJ, Forehand R

(2011). Structure and measurement of depression in youths: applying item response theory to clinical data. *Psychological Assessment* **23**, 819-833.

Coryell W, Winokur G, Shea T, Maser JD, Endicott J, Akiskal HS (1994). The long-term stability of depressive subtypes. *The American Journal of Psychiatry* **151**, 199-204.

Cramer AO, Borsboom D, Aggen SH, Kendler KS (2012). The pathoplasticity of dysphoric episodes: differential impact of stressful life events on the pattern of depressive symptom inter-correlations. *Psychological Medicine* **42**, 957-965.

Cuijpers P, Vogelzangs N, Twisk J, Kleiboer A, Li J, Penninx BW (2014). Comprehensive meta-analysis of excess mortality in depression in the general community versus patients with specific illnesses. *The American Journal of Psychiatry* **171**, 453-462.

Danks D, Fancsali S, Glymour C, Scheines R (2010). Comorbid science? *The Behavioral and Brain Sciences* **33**, 153-155.

Eaton NR, Krueger RF, Markon KE, Keyes KM, Skodol AE, Wall M, Hasin DS, Grant BF (2013). The structure and predictive validity of the internalizing disorders. *Journal of Abnormal Psychology* **122**, 86-92.

Familiar I, Ortiz-Panozo E, Hall B, Vieitez I, Romieu I, Lopez-Ridaura R, Lajous M (2015). Factor structure of the Spanish version of the Patient Health Questionnaire-9 in Mexican women. *International Journal of Methods in Psychiatric Research* **24**, 74-82.

Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJ, Vos T, Whiteford HA (2013). Burden of Depressive Disorders by Country, Sex, Age, and Year: Findings from the Global Burden of Disease Study 2010. *PLoS Medicine* **10**.

Fried EI, Nesse RM (2015). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *Journal of Affective Disorders* **172**, 96-102.

Fried EI, Nesse RM, Guille C, Sen S (2015). The differential influence of life stress on individual symptoms of depression. *Acta Psychiatrica Scandinavica* **131**, 465-471.

Fried EI, Nesse RM, Zivin K, Guille C, Sen S (2014). Depression is more than the sum score of its parts: individual DSM symptoms have different risk factors. *Psychological Medicine* **44**, 2067-2076.

Fried EI, van Borkulo CD, Epskamp S, Schoevers RA, Tuerlinckx F, Borsboom D (2016). Measuring Depression Over Time . . . or not? Lack of Unidimensionality and Longitudinal Measurement Invariance in Four Common Rating Scales of Depression. *Psychological Assessment* Jan 28 [Epub ahead of print]

Goekoop R, Goekoop JG (2014). A network view on psychiatric disorders: network clusters of symptoms as elementary syndromes of psychopathology. *PLoS One* **9**, e112734.

Goldberg DP, Krueger RF, Andrews G, Hobbs MJ (2009). Emotional disorders: cluster 4 of the proposed meta-structure for DSM-V and ICD-11. *Psychological Medicine* **39**, 2043-2059.

Haig BD, Vertue FM (2010). Extending the network perspective on comorbidity. *The Behavioral and Brain Sciences* **33**, 158.

Haslam N, Holland E, Kuppens P (2012). Categories versus dimensions in personality and psychopathology: a quantitative review of taxometric research. *Psychological Medicine* **42**, 903-920.

Humphry SM, McGrane JA (2010). Is there a contradiction between the network and latent variable perspectives? *The Behavioral and Brain Sciences* **33**, 160-161.

Hybels CF, Landerman LR, Blazer DG (2013). Latent subtypes of depression in a community sample of older adults: can depression clusters predict future depression trajectories? *Journal of Psychiatric Research* **47**, 1288-1297.

Jennrich RI, Bentler PM (2011). Exploratory bi-factor analysis. *Psychometrika* **76**, 537-549.

Jennrich RI, Bentler PM (2012). Exploratory bi-factor analysis: The oblique case. *Psychometrika* **77**, 442-454.

Keller MC, Neale MC, Kendler KS (2007). Association of different adverse life events with distinct patterns of depressive symptoms. *The American Journal of Psychiatry* **164**, 1521-1529.

Keller MC, Nesse RM (2006). The evolutionary significance of depressive symptoms: different adverse situations lead to different depressive symptom patterns. *Journal of Personality and Social Psychology* **91**, 316-330.

Kendler KS, Aggen SH, Neale MC (2013). Evidence for multiple genetic factors underlying DSM-IV criteria for major depression. *JAMA Psychiatry* **70**, 599-607.

Kessler RC, Bromet EJ (2013). The epidemiology of depression across cultures. *Annual Review of Public Health* **34**, 119-138.

Kline RB (2011). *Principles and practice of structural equation modelling*. Guilford Press: New York.

Kroenke K, Spitzer RL, Williams JB (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine* **16**, 606-613.

Krueger RF, Deyoung CG, Markon KE (2010). Toward scientifically useful quantitative models of psychopathology: the importance of a comparative approach. *The Behavioral and Brain Sciences* **33**, 163-164.

Lewinsohn PM, Pettit JW, Joiner TE Jr., Seeley JR (2003). The symptomatic expression of major depressive disorder in adolescents and young adults. *Journal of Abnormal Psychology* **112**, 244-252.

Li Y, Aggen S, Shi S, Gao J, Tao M, Zhang K, Wang X, Gao C, Yang L, Liu Y, Li K, Shi J, Wang G, Liu L, Zhang J, Du B, Jiang G, Shen J, Zhang Z, Liang W, Sun J, Hu J, Liu T, Miao G, Meng H, Hu C, Huang G, Li G, Ha B, Deng H, Mei Q, Zhong H, Gao S, Sang H, Zhang Y, Fang X, Yu F, Yang D, Chen Y, Hong X, Wu W, Chen G, Cai M, Song Y, Pan J, Dong J, Pan R, Zhang W, Shen Z, Liu Z, Gu D, Liu X, Zhang Q, Flint J, Kendler KS (2014a). The structure of the symptoms of major depression: exploratory and

confirmatory factor analysis in depressed Han Chinese women. *Psychological Medicine* **44**, 1391-1401.

Li Y, Aggen S, Shi S, Gao J, Tao M, Zhang K, Wang X, Gao C, Yang L, Liu Y, Li K, Shi J, Wang G, Liu L, Zhang J, Du B, Jiang G, Shen J, Zhang Z, Liang W, Sun J, Hu J, Liu T, Miao G, Meng H, Hu C, Huang G, Li G, Ha B, Deng H, Mei Q, Zhong H, Gao S, Sang H, Zhang Y, Fang X, Yu F, Yang D, Chen Y, Hong X, Wu W, Chen G, Cai M, Song Y, Pan J, Dong J, Pan R, Zhang W, Shen Z, Liu Z, Gu D, Liu X, Zhang Q, Flint J, Kendler KS (2014b). Subtypes of major depression: latent class analysis in depressed Han Chinese women. *Psychological Medicine*, **44**, 3275-3288.

Lux V, Kendler KS (2010). Deconstructing major depression: a validation study of the DSM-IV symptomatic criteria. *Psychological Medicine* **40**, 1679-1690.

MacCallum RC, Roznowski M, Necowitz LB (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin* **111**, 490-504.

Markon KE, Chmielewski M, Miller CJ (2011). The reliability and validity of discrete and continuous measures of psychopathology: a quantitative review. *Psychological Bulletin* **137**, 856-879.

Markus KA (2010). Questions about networks, measurement, and causation. *The Behavioral and Brain Sciences* **33**, 164-165.

Melartin T, Leskela U, Rytsala H, Sokero P, Lestela-Mielonen P, Isometsa E (2004). Co-morbidity and stability of melancholic features in DSM-IV major depressive disorder. *Psychological Medicine* **34**, 1443-1452.

Mezuk B, Kendler KS (2012). Examining variation in depressive symptoms over the life course: a latent class analysis. *Psychological Medicine* **42**, 2037-2046.

Molenaar PC (2010). Latent variable models are network models. *The Behavioral and Brain Sciences* **33**, 166.

Murray CJ, Lopez AD (1996). Evidence-based health policy--lessons from the Global Burden of Disease Study. *Science* **274**, 740-743.

Muthén BO (1989). Dichotomous Factor Analysis of Symptom Data. *Sociological Methods and Research* **18**, 19-65.

Olsen LR, Jensen DV, Noerholm V, Martiny K, Bech P (2003). The internal and external validity of the Major Depression Inventory in measuring severity of depressive states. *Psychological Medicine* **33**, 351-356.

Oquendo MA, Barrera A, Ellis SP, Li S, Burke AK, Grunebaum M, Endicott J, Mann JJ (2004). Instability of symptoms in recurrent major depression: a prospective study. *The American Journal of Psychiatry* **161**, 255-261.

Pornprasertmanit S, Miller P, Schoemann A, Rosseel Y (2013). semTools: Useful tools for structural equation modeling. R package version 0.4-0.

Prisciandaro JJ, Roberts JE (2009). A comparison of the predictive abilities of dimensional and categorical models of unipolar depression in the National Comorbidity Survey. *Psychological Medicine* **39**, 1087-1096.

R Core Team (2014). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna.

Reise SP, Moore TM, Haviland MG (2010). Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment* **92**, 544-559.

Revelle W (2013). psych: Procedures for Personality and Psychological Research. Version 1.3.10. Northwestern University, Evanston, Illinois, USA.

Rindskopf D (1984). Structural Equation Models: Empirical Identification, Heywood Cases, and Related Problems. *Sociological Methods and Research* **13**, 109-119.

Rodgers S, Grosse Holtforth M, Muller M, Hengartner MP, Rossler W, Ajdacic-Gross V (2014). Symptom-based subtypes of depression and their psychosocial correlates: a person-

centered approach focusing on the influence of sex. *Journal of Affective Disorders* **156**, 92-103.

Ross D (2010). Some mental disorders are based on networks, others on latent variables. *The Behavioral and Brain Sciences* **33**, 166-167.

Rosseel Y (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software* **48**, 1-36.

Shafer AB (2006). Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *Journal of Clinical Psychology* **62**, 123-146.

Simon GE, Perlis RH (2010). Personalized medicine for depression: can we match patients with treatments? *The American Journal of Psychiatry* **167**, 1445-1455.

Slade T, Andrews G (2005). Latent structure of depression in a community sample: a taxometric analysis. *Psychological Medicine* **35**, 489-497.

Steiger JH (1980). Testing pattern hypotheses on correlation matrices: Alternative statistics and some empirical results. *Multivariate Behavioral Research* **15**, 335-352.

Studer J, Baggio S, Mohler-Kuo M, Dermota P, Gaume J, Bertholet N, Daeppen JB, Gmel G (2013a). Examining non-response bias in substance use research--are late respondents proxies for non-respondents? *Drug and Alcohol Dependence* **132**, 316-323.

Studer J, Mohler-Kuo M, Dermota P, Gaume J, Bertholet N, Eidenbenz C, Daeppen JB, Gmel G (2013b). Need for informed consent in substance use studies--harm of bias? *Journal of Studies on Alcohol and Drugs* **74**, 931-940.

van Borkulo C, Boschloo L, Borsboom D, Penninx BWJH, Waldorp LJ, Schoevers RA (2015). Association of Symptom Network Structure With the Course of Longitudinal Depression. *JAMA Psychiatry* **72**, 1219-1219.

van Borkulo CD, Borsboom D, Epskamp S, Blanken TF, Boschloo L, Schoevers RA, Waldorp LJ (2014). A new method for constructing networks from binary data. *Scientific Reports* **4**, 5918.

van Loo HM, de Jonge P, Romeijn JW, Kessler RC, Schoevers RA (2012). Data-driven subtypes of major depressive disorder: a systematic review. *BMC Medicine* **10**, 156.

World Health Organization ed. (2012). Taschenführer zur ICD-10 Klassifikation psychischer Störungen nach dem Pocket Guide von J.E. Cooper [Pocket guide to ICD-10 classification of mental and behaioural disorders with glossary and diagnostic criteria for research ICD: CDR-10]. Verlag Hans Huber: Bern.

Zimmerman M, Ellison W, Young D, Chelminski I, Dalrymple K (2015). How many different ways do patients meet the diagnostic criteria for major depressive disorder? *Comprehensive Psychiatry* **56**, 29-34.

Table 1: Symptoms of depression in the ICD-10-based WHO-MDI and the DSM-IV-based PHQ-9

Symptoms as assessed in the WHO-MDI (ICD-10)	**Translation rule** [a]	**Symptoms as assessed in the PHQ-9 (DSM-IV)**	**Dichotomization rule** [b]	
			WHO-MDI	*PHQ-9*
depressed mood		depressed mood	"most of the time"	"more than half of the days"
anhedonia		anhedonia	"most of the time"	"more than half of the days"
lack of energy/fatigue		lack of energy/fatigue	"most of the time"	"more than half of the days"
feelings of worthlessness feelings of guilt	} highest score	feelings of worthlessness and guilt	"more than half of the time"	"more than half of the days"
life not worth living		life not worth living	"more than half of the time"	"more than half of the days"
concentration problems		concentration problems	"more than half of the time"	"more than half of the days"
feeling restless feeling subdued or slowed down	} highest score	psychomotor disturbances	"more than half of the time"	"more than half of the days"
sleeping problems		sleeping problems	"more than half of the time"	"more than half of the days"
reduced appetite increased appetite	} highest score	appetite changes	"more than half of the time"	"more than half of the days"

Note. ICD-10: International Classification of Diseases 10th version; DSM-IV: Diagnostic and Statistical Manual of Mental Disorders 4th edition. WHO-MDI: World Health Organization Major Depression Inventory; PHQ: Patient Health Questionnaire.

[a] Translation rule to combine ICD-10 symptoms into DSM-IV symptoms. The rule is to take the highest value of the relevant ICD-10 symptoms to represent the corresponding DSM-IV symptom.

[b] Threshold for scoring the symptom as "present".

Table 2: Baseline characteristics of study participants

C-SURF sample	**Baseline**	*Depression (%)* [a]	**Follow-up**	*Depression (%)* [a]
Total	5883	6.1	5155	7.0
Age (M ±SD)	20.0 ± 1.2	-	21.3 ± 1.2	-
Below median	2940 (50.0%)	5.4	2583 (50.1%)	5.8
Above median	2937 (50.0%)	6.7	2570 (49.9%)	8.3
Education				
Primary school	2842 (48.5%)	5.9	358 (7.0%)	13.4
Secondary vocational education	1870 (31.9%)	6.0	2871 (55.9%)	6.4
Secondary school education	1054 (18.0%)	6.5	1659 (32.3%)	7.1
Above secondary	99 (1.7%)	6.1	248 (4.8%)	4.4
Linguistic region				
German	2658 (45.2%)	5.3	2335 (45.3%)	5.9
French	3225 (54.8%)	6.7	2820 (54.7%)	8.0
NHANES 2005-2012 samples	**Men**	*Depression (%)* [a]	**Women**	*Depression (%)* [a]
Total	2174	4.9 [b]	2244	7.6 [b]
Age (M ±SD)	22.2 ± 3.3	-	22.4 ± 3.3	-
Below median	1218 (56.0%)	5.6 [b]	1195 (53.3%)	7.6 [b]
Above median	956 (44.0%)	4.7 [b]	1049 (46.7%)	8.9 [b]
Education				
Primary school	650 (29.9%)	7.8 [b]	531 (23.7%)	15.1 [b]
Secondary education	607 (27.9%)	6.0 [b]	575 (25.6%)	10.1 [b]
Above secondary	915 (42.1%)	3.8 [b]	1138 (50.7%)	6.2 [b]
Race				
Non-Hispanic white	764 (35.1%)	4.2 [b]	778 (34.7%)	6.9 [b]
Non-Hispanic black	552 (25.4%)	6.7 [b]	548 (24.4%)	14.5 [b]
Mexican American	501 (23.0%)	5.2 [b]	503 (22.4%)	9.6 [b]
Other	357 (16.4%)	8.3 [b]	415 (18.5%)	9.1 [b]

Note. M: Mean. SD: Standard Deviation.

[a] Prevalence of current depression of at least mild-moderate degree

[b] Average prevalence rate calculated across the NHANES cycles 2005-2012. The prevalence rates within each cycle were calculated using weighted data.

Table 3: Summary of tetrachoric correlations of the 9 DSM-IV depression symptoms across general-population samples and subsamples of currently depressed subjects

Sample	**Symptom correlations**		
	M	*IQR*	*Range*
General population			
C-SURF baseline	0.74	0.65-0.78	0.52-0.85
C-SURF follow-up	0.69	0.64-0.78	0.49-0.85
NHANES 2009-2012 men	0.55	0.51-0.60	0.44-0.73
NHANES 2009-2012 women	0.58	0.52-0.61	0.31-0.75
Depressed			
C-SURF baseline	0.24	0.09-0.34	-0.05-0.69
C-SURF follow-up	0.22	0.12-0.38	-0.06-0.65
NHANES 2009-2012 men	0.04	-0.04-0.19	-0.24-0.47
NHANES 2009-2012 women	0.09	-0.03-0.17	-0.34-0.38

Note. M: Median; IQR: Inter-Quartile Range.

Table 4: Comparison of tetrachoric correlations of the 9 DSM-IV depression symptoms in general-population samples versus subsamples of currently depressed subjects

Samples compared		**Steiger test** [a]		**Ratios of squared correlations** [b]		
		χ^2 (df)	*p-value*	*M*	*IQR*	*Range*
C-SURF baseline:	*general-population / depressed*	5840.5 (36)	< 0.0001	8.4	4.9-51.8	1.5-221100.0
C-SURF follow-up:	*general-population / depressed*	4777.6 (36)	< 0.0001	9.4	3.7-28.8	1.5-561.7
NHANES 2009-2012 men:	*general-population / depressed*	1362.7 (36)	< 0.0001	30.9	6.7-128.8	2.5-22460.0
NHANES 2009-2012 women:	*general-population / depressed*	2292.1 (36)	< 0.0001	14.2	8.1-46.2	0.8-2472.0

Note. M: Median; IQR: Inter-Quartile Range. df: degrees of freedom.

[a] Tests the hypothesis that two correlation matrices differ from each other

[b] For each symptom pair, its squared correlation in the general-population sample was divided by its squared correlation in the corresponding sample of depressed. Ratios > 1.0 indicate that the correlation was higher in the general population than among depressed.

Table 5: Summary of exploratory and confirmatory factor and bifactor analyses of the 9 DSM-IV depression symptoms in general-population samples and subsamples of currently depressed subjects

	Model 1 (1-factor CFA)			**Model 2 (EFA)**				**Model 3 (model 1 including residual co-variances)**			**Model 4.a1 (theoretical bifactor model 1)**		
Sample	*CFI*	*TLI*	*RMSEA*	*Number of factors extracted*	*CFI*	*TLI*	*RMSEA*	*CFI*	*TLI*	*RMSEA*	*CFI*	*TLI*	*RMSEA*
General-population													
C-SURF baseline	0.994	0.993	0.026	1	0.994	0.993	0.026	1.000	1.000	0.000	- [b]	- [b]	- [b]
C-SURF follow-up	0.993	0.991	0.027	1	0.993	0.991	0.027	1.000	1.001	0.000	- [b]	- [b]	- [b]
NHANES 2005-2012 men	0.983	0.977	0.029	1	0.983	0.977	0.029	0.994	0.990	0.018	- [b]	- [b]	- [b]
NHANES 2005-2012 women	0.980	0.973	0.039	1	0.980	0.973	0.039	1.000	1.002	0.000	- [b]	- [b]	- [b]
Depressed													
C-SURF baseline	0.867	0.822	0.084	4	0.991	0.944	0.047	0.989	0.978	0.029	- [b]	- [b]	- [b]
C-SURF follow-up	0.925	0.900	0.064	2	0.968	0.939	0.050	0.995	0.992	0.018	- [b]	- [b]	- [b]
NHANES 2005-2012 men	1.00	1.107	0.000	1	1.00	1.107	0.000	1.000	1.107	0.000	- [b]	- [b]	- [b]
NHANES 2005-2012 women	0.556	0.407	0.059	- [a]	- [a]	- [a]	- [a]	0.944	0.916	0.022	- [b]	- [b]	- [b]

	Model 4.a2 (theoretical bifactor model 2)			**Model 4.a3 (theoretical bifactor model 3)**			**Model 4.b (exploratory bifactor model)**		
Sample	*CFI*	*TLI*	*RMSEA*	*CFI*	*TLI*	*RMSEA*	*CFI*	*TLI*	*RMSEA*
General-population									
C-SURF baseline	- [b]	- [b]	- [b]	- [b]	- [b]	- [b]	- [b]	- [b]	- [b]
C-SURF follow-up	- [b]	- [b]	- [b]	- [b]	- [b]	- [b]	- [b]	- [b]	- [b]
NHANES 2005-2012 men	1.000	1.004	0.000	- [b]	- [b]	- [b]	0.989	0.976	0.029
NHANES 2005-2012 women	0.997	0.994	0.019	- [b]	- [b]	- [b]	- [b]	- [b]	- [b]
Depressed									
C-SURF baseline	0.959	0.917	0.057	- [b]	- [b]	- [b]	1.000	1.009	0.000
C-SURF follow-up	- [b]	- [b]	- [b]	0.975	0.951	0.045	- [b]	- [b]	- [b]
NHANES 2005-2012 men	1.000	1.545	0.000	1.000	1.432	0.000	- [b]	- [b]	- [b]
NHANES 2005-2012 women	- [b]	- [b]	- [b]	- [b]	- [b]	- [b]	- [b]	- [b]	- [b]

Note. CFA: Confirmatory Factor Analysis. EFA: Exploratory Factor Analysis. CFI: Comparative Fit Index; TLI: Tucker-Lewis Index; RMSEA: Root Mean Square Error of Approximation

[a] None of the admissible models reached the criteria for good model fit.

[b] Model inadmissible due to negative residual variance of at least one symptom.

Table 6: Correlations of the 9 DSM-IV depression symptoms adjusted for overall depression severity as estimated by confirmatory factor analysis (model 3)

		General-population				Depressed			
Symptom pair		*C-SURF baseline*	*C-SURF follow-up*	*NHANES 2005-2012 men*	*NHANES 2005-2012 women*	*C-SURF baseline*	*C-SURF follow-up*	*NHANES 2005-2012 men*	*NHANES 2005-2012 women*
Depressed Mood	– Anhedonia	**0.40**	**0.32**		0.10				
Depressed Mood	– Fatigue/energy			-0.17	-0.14				
Depressed Mood	– Worthlessness/guilt			**0.26**	0.22	**0.48**			**0.34**
Depressed Mood	– Life not worth		**0.29**				0.24		
Depressed Mood	– Concentration					-0.32			
Depressed Mood	– Sleep Problems				-0.22				
Depressed Mood	– Appetite Changes				**-0.39**				
Anhedonia	– Fatigue/energy	**0.32**	**0.40**		**-0.26**	0.35	**0.39**		
Anhedonia	– Concentration						-0.23		
Anhedonia	– Psychomotor Changes	-0.09				**-0.33**			
Anhedonia	– Sleep Problems	-0.08	-0.10		**-0.27**				
Anhedonia	– Appetite Changes	**-0.11**	-0.07						
Fatigue/energy	– Worthlessness/guilt				-0.17				
Fatigue/energy	– Life not worth	**-0.35**	**-0.29**			-0.15	**-0.31**		
Fatigue/energy	– Sleep Problems			**0.34**	**0.39**				**0.33**
Worthlessness/guilt	– Psychomotor Changes				**-0.35**				**-0.29**
Life not worth	– Concentration					**0.23**			
Life not worth	– Sleep Problems				**-0.31**				
Life not worth	– Appetite Changes				**-0.46**				
Concentration	– Psychomotor Changes	**0.20**				**0.28**	0.26		
Psychomotor Changes	– Sleep Problems	**0.22**				**0.26**			
Psychomotor Changes	–Appetite Changes						**-0.25**		
Sleep Problems	– Appetite Changes	**0.22**	**0.23**			**0.38**	0.17		

Note. Correlations printed in bold are statistically significant with $p < 0.05$.